# ARIF KHAN

+91-9136259776 | arifkhan50636@gmail.com | http://www.linkedin.com/in/arik7 | https://github.com/Arifkhd

## EDUCATION

**Master of Science in Computer Science**                    **2024 – Pursuing (2026)**
Ramniranjan Jhunjhunwala College, Mumbai

**Bachelor of Science in Computer Science**                    **2021-2024**
Ramniranjan Jhunjhunwala College, Mumbai                    8.28 CGPA

## TECHNICAL SKILLS

**Programming & Data Handling:** Python, SQL, R Programming

**Frameworks & Libraries:** Pandas, NumPy, Keras, Scikit-learn, TensorFlow, Matplotlib, Seaborn

**Data Analytics & Visualization:** Excel, Tableau, Power BI, Data Visualization

**Machine Learning & AI:** Machine Learning, Deep Learning, Artificial Intelligence, MLOps, Large Language Models (LLM), Natural Language Processing (NLP)

**Statistical & Analytical Methods:** Statistics, Time Series Analysis, Forecasting Analysis, Data Analytics

**Version Control & Workflow:** Git

**Soft Skills:** Critical Thinking, Problem Solving, Business Acumen, Analytical Thinking, Research

## CERTIFICATE

Data Science Certification, ExcelR Solutions
Generative AI with Hugging face and Langchain, Udemy

## EXPERIENCE

**Data Scientist Intern | Ai Variant, Hyderabad, India**                    **Mar 24 – Mar 25**

- I developed predictive models using Linear Regression, Random Forest, and XGBoost for regression tasks. For classification, I implemented Logistic Regression, SVM, and Neural Networks.
- In NLP, I worked on text classification and sentiment analysis using TF-IDF, Word Embeddings, and Transformer models. This internship enhanced my skills in data preprocessing, model evaluation, and deployment with Python, Scikit-learn, and TensorFlow.

## PROJECTS

**House Price Prediction**                                        **Live Demo**

- The Bangalore House Price Prediction project aims to build a machine learning model to estimate house prices based on factors like location, size, and amenities.
- The project involves data preprocessing, feature engineering, and model training using algorithms such as Linear Regression, Random Forest, and XGBoost.
- Evaluation metrics like RMSE and R² Score are used to assess model performance. Finally, a web application is deployed using Streamlit for real-time predictions.

**Drug Prediction Using NLP**                                        **Live Demo**

- Developed an NLP-based machine learning model to classify medical conditions like Depression, High Blood Pressure, and Type 2 Diabetes from patient reviews.
- Conducted data preprocessing, including tokenization, stop-word removal, and TF-IDF vectorization.
- Implemented and evaluated models using techniques like Logistic Regression and Random Forest, achieving high accuracy and meaningful insights into drug effectiveness.

**Car Price Prediction**                                        **Live Demo**

- The Car Price Prediction project aims to estimate the selling price of cars based on features like brand, model, year, seller type, kilo meter, owner type, fuel type, and transmission.
- It involves data preprocessing, feature engineering, and model training using algorithms such as Linear Regression, Random Forest, and XGBoost.
- Evaluation metrics like RMSE and R² Score are used to assess model accuracy. A web application is deployed using Streamlit for real-time price predictions.